STAT 401 Chapter 4 except 4.6

Zepu Zhang
October 22, 2010

1 Simultaneous inference

1.1 Need for "simultaneous intervals"

Suppose we want to make estimation/prediction for multiple quantities and construct a confidence interval (CI) for each. Sometimes we want to make a statistical statement like this: with confidence coefficient 0.95, all CI's contain the true values of their respective target quantities (that is, 95% of the time the recipe works as a whole).

If each CI is constructed as a 0.95 CI for its quantity of interest independent of the other CI's, the pack of CI's won't simultaneously contain the truth 95% of the time—the confidence coefficient will be below 0.95. It would be 0.95 only if all CI's <u>always</u> work or fail simultaneously, and this is never the case.

Therefore some special procedure is needed. One can imagine the individual CI's need to have confidence coefficients higher than 0.95 in order for the whole pack to have a confidence coefficient of 0.95. For example, each CI "works" 98% of the time and then it might be possible that they simultaneously "work" as a whole 95% of the time.

1.2 The Bonferroni procedure

Suppose we construct a CI for k quantities, with confidence coefficients $1-\alpha_1$,..., $1-\alpha_k$. Let A_i, $i = 1, \ldots, k$, be the event that the ith CI contains the true value of the ith quantity, then $P(A_i) = 1 - \alpha_i$. Our goal is to make $P(A_1 \cap A_2 \cap \cdots \cap A_k) = 1 - \alpha$, say 0.95. This says the intervals <u>simultaneously</u> contain the respective true values with probability 0.95. Since

$$\overline{A_1 \cap \cdots \cap A_k} = \overline{A_1} \cup \cdots \cup \overline{A_k}$$

(the two sides are succinctly written as $\overline{\cap_i A_i}$ and $\cup_i \overline{A_i}$, respectively), we have

$$P(\cap_i A_i) = 1 - P\left(\overline{\cap_i A_i}\right) = 1 - P\left(\cup_i \overline{A_i}\right)$$

Noticing

$$P\left(\cup_i \overline{A_i}\right) \le P\left(\overline{A_1}\right) + \cdots + P\left(\overline{A_k}\right) = \alpha_1 + \cdots + \alpha_k$$

that is,

$$P(\cap_i A_i) \ge 1 - \sum_i \alpha_i$$

we see $P(\cap_i A_i) \ge 1 - \alpha$ will be achieved if we make

$$1 - \sum_i \alpha_i = 1 - \alpha, \quad \text{i.e.} \quad \sum_i \alpha_i = \alpha$$

Therefore if we require a $1 - \alpha$ simultaneous confidence coefficient, we can guarantee this by choosing the α_i's such that $\sum_i \alpha_i = \alpha$. If we choose identical α_i's then each is α/k.

This is a <u>conservative</u> procedure: the actual confidence coefficient achieved is likely higher than $1 - \alpha$ (but its exact value is unknown) due to the inequality in the argument above.

This is most useful when k is not big. For otherwise the individual α_i's may become too small and the individual CI's too wide to be useful.

Bonferroni is a <u>general</u> procedure. For specific questions there may exist more fine-tuned procedures that give narrower intervals which still achieve the required simultaneous confidence coefficient.

1.3   Joint estimation of β_0 and β_1

Construct simultaneous $1 - \alpha$ CI's for β_0 and β_1:

$$b_0 \pm B s\{\hat{\beta}_0\}, \quad b_1 \pm B s\{\hat{\beta}_1\}$$

where

$$B = t(1 - \alpha/4, n - 2)$$

(The individual CI's will have confidence coefficient $1 - \alpha/2$, hence the critical value is $t(1 - \alpha/4, n - 2)$.)

Comment 5 on page 157.

Generalization   It is straightforward to generalize this idea to multiple regression coefficients in a multiple regression model.

Example   Page 156.

1.4   Simultaneous estimation of $E[Y]$

For X levels $x_1, \ldots, x_k$, get

$$\hat{Y}_i = b_0 + b_1 x_i$$

and
$$s^2\{\hat{Y}_i\} = s^2\big(1/n + (x_i - \bar{x})^2/S_{xx}\big).$$

Then build individual CI's with the same multiplier
$B = t(1 - \alpha/2k, n - 2)$:

$$\hat{Y}_i \pm Bs\{\hat{Y}_i\}$$

Note   When we use matrix notation (as we should), we will not write and obtain the sample variance for each $\hat{Y}_i$. Instead, they are the diagonal elements of the covariance matrix of $\vec{\hat{Y}}$.

1.5   Simultaneous prediction intervals for new observations

Similar to above. Just need a $1 - \alpha/k$ interval for each.

2   Regression through origin

Sometimes the nature of the problem dictates $E(Y) = 0$ when $X = 0$. In such situations we should force $\beta_0 = 0$, i.e. the model is

$$Y = \beta_1 X + \epsilon.$$

Assumptions about ϵ are as before.

In multiple regression, the model is

$$\vec{Y} = \mathbf{X}\vec{\beta} + \vec{\epsilon}$$

where the design matrix $\mathbf{X}$ does not have a leading column of 1's.

2.1   Inference of the model parameters

In matrix notation, notice this changed design matrix does not change anything in our derivation of the LS estimator, which is still

$$\vec{\hat{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\vec{Y}$$

which is normal with mean $\vec{\beta}$ (if the true model indeed has zero intercept) and covariance matrix $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.

If it is a SLR (i.e. one predictor only), then $\vec{\beta}$ has only one element and $\mathbf{X}$ has only one column. It is easy to see the estimator is

$$\hat{\beta}_1 = \frac{\sum X_i Y_i}{\sum X_i^2}$$

Its mean is β_1 and variance is $\sigma^2 / \sum_i X_i^2$.

An unbiased estimator of σ^2 is again

$$S^2 = \text{MSE} = \frac{\sum e_i^2}{n-p}$$

Note the numerator in a SRL will be $n-1$, because p is 1 in that case (only one parameter, β_1, is estimated from the data). You can also think the number of predictors is now the number of X's; there is not that extra predictor 1 (which would be paired with its coefficient β_0).

In a SLR, the un-restrained regression line goes through $(\overline{X}, \overline{Y})$ whereas the restrained regression line goes through $(0,0)$. I suppose a similar statement can be made for MLR but I did not check.

Example   Page 162.

R tip   To enforce a zero intercept, added a -1 predictor in the formula, like, `lm(y ~ -1 + x1 + x2, data)`.

2.2   Further inferences

Using matrix notation, indeed all the formulas remain intact. Only remember that the design matrix now does not have the leading column of 1's and the predictor vector ($\vec{x}$) does not have a leading element of 1. Also notice that the df of the t distribution is $n-p$, where p does not count an extra constant predictor.

3   Effects of measurement errors

3.1   Measurement errors in Y

If the error is independent of X and Y and, better yet, has a normal distribution with constant variance that is not affected by X, this error will be simply absorbed by the ϵ in the model. In this case, estimation of $\vec{\beta}$ stay unchanged, and their nice properties (unbiased; min var among unbiased linear estimators) still hold.

3.2   Measurement errors in X

This is trickier and we don't worry about a systematic treatment.

When the data X are corrupted, they are not the "true" X values. Estimates of $\vec{\beta}$ using these X ignoring the possible errors will introduce bias.

4   Choice of X levels

Some empirical guidelines in section 4.7, page 171.

If the X levels are under our control (e.g. in experimental settings), how to choose the X levels depends on the goal of the study. The motivation is to have an estimator (of the thing of interest) that is unbiased and has a small sampling variance.